

## DOCUMENT RESUME

ED 250 395

TM 840 749

AUTHOR Scheuneman, Janice Dowd  
TITLE The Female Voice in Measurement and Methodology.  
PUB DATE Apr 84  
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).  
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)  
EDRS PRICE MF01/PC01 Plus Postage.  
DESCRIPTORS \*Females; \*Males; Measurement Techniques; Moral Issues; \*Research Methodology; Statistics; Test Bias

## ABSTRACT

In this paper, a parallel is drawn between the sex differences in moral reasoning suggested by Gilligan (1982) and a masculine and a feminine approach to measurement and statistics. A masculine approach could be seen to focus on the application of the rules of mathematics and statistics to measurement problems in ways which rational observers would agree are appropriate. The female approach would be to focus on the limitations of the rules and the context in which they are applied. Examples from the areas of test and items bias are provided. The conclusion is that both modes of thought are necessary for the performance of sound research.  
(Author)

\*\*\*\*\*  
\* Reproductions supplied by EDRS are the best that can be made \*  
\* from the original document. \*  
\*\*\*\*\*

ED250395

U.S. DEPARTMENT OF EDUCATION  
NATIONAL INSTITUTE OF EDUCATION  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

X This document has been reproduced as  
received from the person or organization  
originating it.  
Minor changes have been made to improve  
reproduction quality.

• Points of view or opinions stated in this docu-  
ment do not necessarily represent official NIE  
position or policy.

The Female Voice in Measurement  
and Methodology

Janice Dowd Scheuneman  
Educational Testing Service

"PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

J. Scheuneman

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)."

Presented in a symposium sponsored by the Committee on the Role and Status  
of Women in Educational Research, H. S. Farmer (Chair), Integrating the New  
Scholarship on Women into Required Courses in Schools of Education.  
American Educational Research Association, New Orleans, April 1984

## Abstract

In this paper, a parallel is drawn between the sex differences in moral reasoning suggested by Gilligan (1982) and a masculine and a feminine approach to measurement and statistics. A masculine approach could be seen to focus on the application of the rules of mathematics and statistics to measurement problems in ways which rational observers would agree are appropriate. The female approach would be to focus on the limitations of the rules and the context in which they are applied. Examples from the areas of test and item bias are provided. The conclusion is that both modes of thought are necessary for the performance of sound research.

When I bought Carol Gilligan's book, In a Different Voice (1982), I did not expect it to be applicable to my work in measurement. I was surprised and excited, therefore, that the contrast she drew between the male and female perspectives and modes of thinking could stand as a model for explaining some of the conflict within my own area of specialization. In this paper I will elaborate some of these connections and suggest how our female voice is related to our research, even in the "objective" areas of educational measurement and statistics.

Before beginning I would like to quote a caveat Gilligan made concerning her work and apply this my own remarks as well.

The different voice that I describe is characterized not by gender but theme. Its association with women is an empirical observation, and it is primarily through women's voices that I trace its development. But the association is not absolute, and the contrasts between male and female voices are presented here to highlight a distinction between two modes of thought and to focus on a problem of interpretation rather than to present a generalization about either sex. (p.2)

In her book, Gilligan discusses the findings of Piaget with regard to children's games. In their games, boys become "increasingly fascinated with the legal elaboration of rules and the development of fair procedures for adjudicating conflicts." (Gilligan, p.10) From this perspective, one could see whole areas of adult pursuit as further elaborated childhood games for which there are important payoffs. Mathematics is an example where the rules define the pursuit almost in entirety and the quantitative aspects of statistics and psychometrics are bound up in these mathematical rules.

In general we as researchers, and in particular we who have chosen to specialize in quantitative areas, have become adept at the rules of our respective versions of this adult game. The very fact that we have succeeded in our training and in mastering our discipline reflect our adaptability to this masculine mode of thought. Piaget goes on to point out, however, that girls have "a more 'pragmatic' attitude toward rules, 'regarding a rule good as long as the game repaid it.'" (cited in Gilligan, p.10) I believe that many of us have almost unconsciously brought this "pragmatic" attitude toward rules into our work giving us a perspective which may enhance our science but may also be ignored or rejected by the masculine establishment for whom the rules are paramount.

I would like to illustrate the influence of the female voice within my own field of specialization, test and item bias, where I am most familiar with the literature. Gilligan's work is also particularly applicable here since the issues of test bias are moral issues. How fair are our tests, our selection procedures? Are we preventing access to a better life for persons from minority cultures, particularly those with economic disadvantage. Instances of the impact of the female perspective most likely exist within other areas of endeavor, but because of their moral quality, these issues serve as good examples of the differences in approach under discussion here.

Gilligan distinguishes between the higher levels of Kohlberg's moral development stages and the more typical female response as follows:

Whereas the rights conception of morality that informs Kohlberg's principled level (stages five and six) is geared to arriving at an objectively fair or just resolution to moral dilemmas upon which all rational persons could agree, the

responsibility conception focuses instead on the limitations of any particular resolution and describes the conflicts that remain.

In counterpoint to this distinction is a discussion of Nancy Cole's in contrasting the two major factions involved in bias research.

Many people motivated by Messick's ethical questions of social policy began to pursue scholarly research on test bias at the technical validity level. It is interesting that these policy-level motivations were generally of two opposing types. On the one hand, much test bias research has been motivated by the belief that tests are socially constructive and by the desire to defend them from attack. On the other hand, perhaps as much research has been motivated by a concern that tests might be contributing, through bias, to a harmful situation and the desire to find any bias that exists and to eliminate it. Thus, many researchers in the test bias area can be roughly characterized as belonging to one of two groups: the defenders or the reformers. The defenders expected to find no bias and to contribute to the solution of several social issues by eliminating the specter of bias from the public debates. The reformers expected to find bias and to contribute to the solution of the same issues by forcing reforms to correct bias.

The defenders have tended to use established measurement principles and procedures to demonstrate their contention that the tests are not biased, an objective resolution "upon which rational persons could agree," but this agreement has not followed. Indeed, some of the writing of the defenders reflects a kind of bewilderment that this should be so.

Jensen in his book, bias in Mental Testing, attempted to distinguish between fairness and bias when he stated, "The assessment of bias is a purely objective, empirical, statistical, and quantitative matter entirely independent of subjective value judgements and ethical issues concerning fairness or unfairness of tests and the uses to which they are put." (p. 375) In this argument he attempts to establish that here is a domain to which we can properly apply our statistical rules. His overemphasis on objectivity betrays a kind of annoyance that the empirical evidence is not being accepted as he thinks it should be. Although he notes that different procedures can give opposing results, he seems unaware of the values inherent in choosing a method most likely to yield the results expected according to those values.

Similarly, in an article defending the sex fairness of Holland's Vocational Preference Inventory, Gottfredson and Holland (1978) state as one of their recommendations, the need to "develop a healthy skepticism about untested intuition. Too often, participants in this controversy have acted as if untested intuitions were equivalent to the results of a program of research." (p.50) While our intuitions should indeed be properly evaluated, one detects a certain testiness which one might suspect comes from frustration that their application of the rules of sound measurement practice have been insufficient to convince "all rational persons." The obvious conclusion seems to be that those who are unconvinced are not behaving rationally.

The reformers, on the other hand, have tended to follow the more female mode of focusing on the limitations. Working from the assumption that bias is present in tests they have looked for new methodologies which will detect the presence of bias in individual test items. It is interesting to note that, in fact, a majority of the researchers who have been pursuing this direction over time have been women. Ron Berk's Handbook of Methods for

Detecting Test Bias, (1982) based on papers delivered at a Johns Hopkins State-of-the-Art Symposium bringing together leading researchers on this topic, shows six of the eight substantive chapters written by women. This is particularly noteworthy as measurement, like many quantitative fields, has a relatively low proportion of women compared to other areas of educational research.

As another example, Wild and Dwyer (1980) illustrate the female tendency to provide context in their critique of methods used to evaluate test bias in an educational admissions setting. The usual design consists of one or more predictors (usually test scores and perhaps grade point average) and a criterion (most often grades at the end of the first year of study). The various procedures are generally dependent on the assumption that the criterion is unbiased, an assumption most concede is unrealistic.

In response Wild and Dwyer have developed an evaluation model in which the first step is the specification of the ultimate criteria, i.e., those which are essential to satisfactory performance and which may be difficult to measure directly." (p.159) In picking the intermediate criteria, which can be measured more easily and directly, the first step would be to evaluate the sufficiency of these criteria in measuring the ultimate goals of the program of study. Sufficiency would be determined on the basis of the completeness with which important elements of these ultimate goals are represented. Predictors should then be chosen in terms of their logical relationship to the ultimate criteria and the sufficiency of the predictors evaluated as well. Wild and Dwyer speculate that with properly chosen predictors and criterion measures, selection bias would be reduced. In essence their recommendation is that the selection procedure cannot be properly

evaluated without the context set by the specification of ultimate goals. Their criticism then is not so much that the rules are wrong, but that the elements to which they have been applied have not been adequately specified.

A related issue is evaluating the sufficiency of a statistical procedure for addressing the research question of interest. My sense is that statisticians tend to get carried away with the elegance of their procedures, and when these are appropriate for the analysis of a given data set, do not always evaluate how well these procedures address the central issue of the research. An elaborate, powerful technique providing useful information somewhat related to the question of interest seems generally to be preferred to a simpler technique, perhaps with a larger error component and without a good significance test, which directly addresses the research question.

I do not wish to suggest that elaborate procedures are never valuable. In fact, the female voice seems to be inherently multivariate. I believe, however, that the female mode of thinking is less apt to lead us to the errors described as follows by Samuel Shye (1978):

Unfortunately, social and psychological investigators, in their attempt to attain the rigor of the physical sciences, often resort prematurely to quantitative trimmings in their studies while neglecting another feature essential to all empirical science: formulating a reliable definitional basis for carrying out observations. Thus elaborate statistical methods are often used to analyze observations whose (usually unexplicated) definitional basis is extremely weak. (p.4)

I believe that our female mode of thought which looks to imbed phenomena in context, which focuses on the limitations and shortcomings of our procedures, has often been combined with our masculine mode of objective

application of rules to produce better research. The New Scholarship has focused our attention on these different modes of thought which, once identified, could be fostered and systematized.

In the teaching of measurement and statistics, attention could be drawn to those parts of research where the female voice might be heard most effectively--probably in the formulation of the problem, the selection of appropriate methodology to address the problem, and in a critical appraisal of the results. In measurement, the issues surrounding test bias have been mentioned as an area where the female voice is probably already well represented. Interest measurement and the selection and evaluation of validation criteria, both for school admissions and for job selection and placement, probably need more attention from this perspective. I'm sure there are others.

As in the personal realm, where incorporation of the masculine and feminine aspects of our being make us whole persons, the integration of both the female voice and the male voice will lead us toward the goal of better research and better science.

# References

Berk, R. A. (Ed.). (1982). Handbook of methods for detecting test bias.

Baltimore: Johns Hopkins University Press.

Cole, N. S. (1981). Bias in Testing. American Psychologist, 36, 1067-1077.

Gilligan, C. (1982). In a different voice. Cambridge: Harvard University Press.

Gottfredson, G. D., & Holland, J. L. (1978). Toward beneficial resolution of the interest inventory controversy. In C. K. Tittle & D. G. Zytowski (Eds.), Sex-fair interest measurement: Research and implications. Washington: The National Institute of Education.

Jensen, A. R. (1980). Bias in mental testing. New York: The Free Press.

Shye, S. (1978). On the search for laws in the behavioral sciences. In S. Shye (Ed.), Theory construction and data analysis in the behavioral sciences. San Francisco: Jossey-Bass.

Wild, C. L., & Dwyer, C. A. (1980). Sex bias in selection. In L. J. Th. van der Kamp, W. F. Langerak, & D. N. M. de Gruijter (Eds.), Psychometrics for educational debates. London: John Wiley.